





Introduction

2024 Nobel The Prize in Economics that has just been awarded to Daron Acemoglu, Simon Johnson and James Robinson reflects the growing interest that economists have towards the analysis of the longrun determinants of culture and institutions. An important strand of this literature regards whether culture and institutions persist over time. A major analytical and conceptual challenge for this research program arises from the many contemporary factors that can confound current the variable of interest (e.g., comparative development outcomes). To give an example, imagine that theory suggests that variability in weather conditions that our ancestors experienced contributed to shape preferences in terms of, say, our loss aversion (Galor and Savitskiy, 2018), or in terms of the importance people attach tradition (Giuliano and to Nunn, 2021). If variability in weather conditions in a given location is roughly constant over time, then it would be difficult to separate the effect of past weather variability from the effect of current variability when analyzing the preferences of a population residing in the same area as their ancestors (at least as long as farming is still a major economic activity nowadays).

International migrants offer a uniqueopportunitytoovercome this analytical challenge. Indeed, their movement across political borders allows to make a distinction between the conditions they experience and those experienced by their ancestors, thus weakening concerns the the about correlation between past and

current weather variability. Furthermore, if we consider different groups of migrants (i.e., from different origin countries) who moved to a given destination country, then we can compare individuals who experience the same conditions where they live today but whose origins differ in terms of ancestral climatic More variability. generally, migrants bring with them the potential influence of the distant past, but they are no longer exposed to the current conditions of their country of origin.

This opportunity to disentangle the effect of past v. current factors has been exploited by economists in the last twenty years in a way that Raquel Fernández (2011) called the "epidemiological approach". To the same extent that geneticists



try to isolate the effects of genes by comparing individuals with genetic differences living in the same environment, economists try to identify the effects of factors from the distant past by comparing individuals of different origins residing in the same country. Unsurprisingly, the United States have become a focal point in economic studies that adopted the epidemiological approach as their long history of migration

gives the possibility to compare migrants (and their descendants) coming from a very large set of different origin countries.

Defining the one's own origin through foreign ancestry

Defining individuals' origin is an empirical challenge. Early papers in this approach focused on first-generation immigrants, i.e., foreign-born individuals (e.g., Antecol, 2000; Luttmer and Singhal, 2011), or secondgeneration immigrants, i.e., natives foreign-born with Fernández, parent(s) (e.g., 2007, Giuliano, 2007, Galor and Savitskiy, 2018, Galor et al., 2020), but a large and fastgrowing set of papers uses a different approach to define one's own origin, which draws on a question that has been included by the Census Bureau

since 1980 in the population census and in the American Community Survey. Notably, the questionnaire of the 1980 Census asked: "What is this person's ancestry?", a phrasing that was slightly modified in 1990 ("What is this person's ancestry or ethnic origin?"), and that has remained unchanged since then. The majority of the native-born population reports a foreign ancestry that directly relates to a country, e.g., German, Irish, Italian, or that can be easily related to a country, e.g., French-Canadian, Welsh, Sicilian, and this allows,

then, to associate a native to the variable of interest that has been measured in his or her foreign country of ancestry. When more than a foreign ancestry is reported, only the first one is retained, while the others are discarded.

If we focus on the 2000 Census, the natives report ancestries that can be connected to 109 distinct foreign countries. *Figure 1* reports the main ancestry for the native population in each Public Use Microdata Area (PUMA).



Data sources: Authors' elaboration on the 2000 Census (Ruggles et al., 2023). Notes: For each PUMA, we identify and assign the most prevalent ancestral heritage among native residents.

www.parisschoolofeconomics.eu



A promising approach, but with some potential shortcomings

The use of the question on the ancestry or ethnic origin is attractive for several reasons. First, the question on the parental countries of birth were last included, together with the question on the maternal foreign-born tongue for individuals, in the 1970 census. Second, the number of nativeborn individuals with a foreign ancestry is much larger than the number of first-generation immigrants, thus providing the researcher with a much greater variability. Third, the time passed since ancestral migration substantially mitigates the concerns about the possibility that immigrants' non-random selection could confound the identification. as this has been progressively diluted across generations.

paid little attention to some potential shortcomings of this variable, with the notable exception of the concerns expressed by Galor et al. (2016, 2020), that the paper by Bertoli et al. (2024)describes and documents. The main shortcomings relate the subjective to nature of the information that is elicited by the question on ancestry, and to the potential unobserved heterogeneity in the time passed since ancestral migration.

The first shortcoming can be summarized with a single figure: 44.7 percent. This is the share of the native-born population in the 2000 Census that cannot be connected to a foreign country. It is true, as we have written above, that the majority of the native-born population reports an ancestry that can be connected to a foreign country, but this a rather tiny majority. is 44.7 percent of the natives opt for (i) not reporting any ancestry, (ii) reporting a native ancestry (e.g., American, Afro-American, Native American), or (iii) report a supranational ancestry (e.g., Latin American, European) that cannot be connected to a foreign country. The second shortcoming arises from the simple fact that migratory waves by country of origin are staggered. Therefore, some natives have a more recent history of immigration than others and because time spent at destination directly shapes cultural traits (see. for instance, Giavazzi et al., 2019), it is likely correlated with the (often culture-related) outcomes of interest.

The literature has, however,

A selected sample

The natives that can be used the analysis in represent a subsample of the native population, which is potentially self-selected. Self-selection might arise from the fact that the question on ancestry elicits a subjective answer, which is intimately related to the choice of one's own identity, rather than an objective information, e.g., the paternal and the maternal countries of birth. In particular, one could be legitimately concerned that, if we consider natives with

one or more ancestors from a given foreign country, those choosing to report a foreign ancestry are more strongly attached to the culture of their ancestral country than those that decide not to report that Furthermore, ancestry. the extent of self-selection might vary across various origins, being stronger for ancestral origins towards which natives have a predominantly negative as adopting attitude, that identity is more "costly".

It is, in general, impossible to

demonstrate the empirical relevance of this concern. as we do not have information about the origin of one's own ancestors, except in one case, which is particularly important, as it relates to the individuals of Mexican ancestry. The questionnaire of the 2000 Census also includes the following question: "ls this person Spanish/Hispanic/ Latino?", with one of the four pre-coded answers being "Yes, Mexican, Mexican American, Chicano". This gives us the



opportunity to focus on the native-born that report to be Hispanic (Mexican), and to (i) measure the share of them that do not report a Mexican ancestry, and then (ii) compare some relevant characteristics of those reporting a Mexican ancestry, with those not reporting this ancestry. Interestingly, 27.3 percent of

almost 600k Hispanic the (Mexican) the natives in sample do not report a Mexican ancestry. Among the Hispanic (Mexican) who do not report a Mexican ancestry, 44.3 percent speaks Spanish at home, while the corresponding share for the Hispanic (Mexican) natives, who report a Mexican ancestry, stands at 56.6 percent, i.e.,

(56.6-44.3)/44.3 = 27.8 percent higher. The difference would be presumably even larger for the natives who also chose not to self-identify themselves as Hispanics. Thus, reporting a foreign ancestry is, at least in the case of Mexico, positively correlated with a credible measure of the attachment to the ancestral culture.

Time elapsed since ancestral migration

А second major concern regarding the epidemiological approach relates to an unobservable characteristic of natives reporting foreign ancestry: the time elapsed since their ancestors migrated to the United States. The analysis effectively compares individuals whose families have resided in the United States for centuries (for instance, those reporting German ancestry) to individuals whose ancestors arrived more recently. The average origin-specific time since ancestral migration is clearly unobserved, and it can be proxied with the share of the population reporting a given ancestry that is born in the ancestral country (Figure 2). This is share is correlated with the average year of arrival of immigrants from a more limited set of countries, that we can build using population censuses from 1850 to 1970. If time since ancestral migration is correlated with the dependent independent variables and

of interest, then regressions

endogeneity.

from

Given that most independent variables used in this literature (particularly geographical and climatological factors) vary smoothly over space, combined with the historical pattern of staggered migration flows to the United States, this concern becomes particularly salient. For example, if adopting the norms and cultural values of the destination country takes time, and if natives whose ancestors migrated to the United States a longer time ago had cultural values closer to those of the United States, then unobserved heterogeneity in the time since ancestral migration could undulv magnify differences across individuals with different foreign ancestries.

related concern stems А from the network effects in migrant settlement patterns. Initially, migrants tend to cluster in locations with preexisting communities of the same origin, only dispersing more widely as integration progresses over generations. This creates differential spatial concentrations based on time since ancestral migration, which must be accounted for given that different US states different can have vastlv institutional frameworks and incentive systems. Moreover, these concentrated ethnic communities can affect the persistence of cultural traits - for instance, the incentives maintain ancestral to languages vary significantly with community size and concentration. A clear example is the differential value of using the Spanish language in states like California versus Vermont - the much larger Spanish-speaking community in California creates stronger economic and social incentives maintenance, language for regardless of time since migration. Therefore, when studying cultural outcomes, researchers should consider not only migrants from the same origin country but also those from countries sharing similar cultural characteristics.

suffer



Figure 2. Share of individuals born in the ancestral country



Data sources: Authors' elaboration on the 2000 Census (Ruggles et al., 2023).

Theoretical model

In a model with two foreign countries of origin and two generations of migrants (second and third), it suffices to assume that third-generation migrants report their ancestry less frequently than secondgeneration migrants, and that the propensities to declare an ancestry and to adopt the destination country's norms are not orthogonal, to show that the proportion of third-generation migrants (i.e., basically the time since

Two illustrative examples

The significance of the concerns outlined above is evaluated through an examination of two prominent studies within the literature.

The first is Fernández and Fogli (2006), a pioneering article in this literature. This ancestral immigration) biases the estimated coefficient in the standard epidemiological regression.

Roughly speaking, if thirdgeneration migrants who adopt the cultural traits of the destination country also tend to not declare any foreign ancestry, one will overestimate the extent of actual persistence among third-generation immigrant. This occurs because those who assimilated the destination culture the most do not report any foreign ancestry anymore. Furthermore. the greater the cultural distance between the origin destination countries. and the stronger the correlation between the propensity to report an ancestry and to adopt the destination country's cultural traits, heightening the concern of self-selection into foreign ancestry.

paper explores the cultural persistence of fertility behaviors. More specifically, it explains fertility outcomes among native-born women in the United States with a foreign ancestry using the total fertility rate from 1950 in each corresponding ancestral country as a predictor. Bertoli et al. (2024) extend this analysis by categorizing migrant women by generational status (second, third, fourth generation, or beyond), something that is possible with the data from the General Social Survey, revealing a clear pattern: for 9



out of 14 ancestral countries in the sample the highest number of women are (at least) fourthgeneration immigrants. For four countries-Finland, Italy, Russia, and Sweden-the third generation forms the largest group, while only in the case of Mexico do second-generation constitute immigrants the largest proportion. This distributionaligns with historical patterns of immigration to the United States. A replication of Fernandez and Fogli (2006) allowing for a differential impact of origin-specific fertility by migratory generation reveals that origin norms do not persist fourth-generation among migrants which constitute 68.9 percent of the estimation sample.

The second paper examined is a recent influential study by Giuliano and Nunn (2021). This

paper tests a hypothesis from anthropology evolutionary explaining that societies that have historically experienced climatic variability greater exhibit lower persistence of cultural traits. To do so, it adopts the epidemiological approach by examining of the influence of historical climatic variability in the ancestral country of origin of natives living in the United States, on their propensity report speaking the ancestral language (i.e., a proxy of attachment to tradition). Bertoli et al. (2024) show that the simple addition of five continent of origin fixed effect in the epidemiological regression - a rather crude but straightforward way of accounting for heterogeneous time since immigration which varies mainly across continents - reduces the size of the estimated coefficient by

half, suggesting an upwards bias when heterogeneity in time since ancestral migration is not properly accounted for. Notably, natives reporting an ancestry in a Latin American country might be more likely to speak Spanish at home not only because these countries are associated with a lower historical climatic variability, but also because a much larger fraction of their parents was born in the ancestral country (Figure 2), and because they grew up in locations within the United States with a large share of Spanish-speaking Hispanic individuals (Figure 1).

Concluding remarks

of self-reported The use ancestry is widely considered foundational for studies employing an epidemiological approach. Available since the 1980 Census, self-reported ancestry data mitigate the well-known issue of attenuation bias. However, this approach introduces challenges, including (i) the subjective nature of self-

reported ancestry declarations, and (ii) the comparison of individuals whose ancestors immigrated to the United States at substantially different times. As demonstrated by Bertoli *et al.* (2024), these limitations present significant implications for econometric analysis, and limit what we can learn about the long-term determinants of economic development from the analysis of natives of foreign ancestry in the United States. The diagnostic tools proposed in their study offer researchers a means to evaluate the impact and extent of these limitations.



References

Antecol H., 2000, "<u>An examination of cross-country differences in the gender gap in labor force</u> participation rates", Labour Economics, 7, 409–426.

Bertoli S. et al., 2024, <u>Migration and the Epidemiological Approach: Time and Self-Selection into</u> <u>Foreign Ancestries Matter</u>, IZA Working Paper No. 17356, Institute for the Study of Labor.

Fernández R., 2011, "<u>Does culture matter?</u>" in Handbook of Social Economics, Benhabib J., Bisin A. & Jackson M.O. (eds.), Elsevier, Vol. 1, 481–510.

Fernández R., 2007, "<u>Women, Work, and Culture</u>", Journal of the European Economic Association, 5, 305–332.

Galor O. & Savitskiy V., 2018, Climatic Roots of Loss Aversion, NBER Working Paper 25273.

Galor O., Özak Ö. & Sarid A., 2020, <u>Linguistic Traits and Human Capital Formation</u>, American Economic Association Papers and Proceedings, 110, 309–13.

Galor O., Özak Ö. & Sarid A., 2016, <u>Geographical Origins and Economic Consequences of Language</u> <u>Structures</u>, IZA Working Paper No. 10379, Institute for the Study of Labor.

Giavazzi F., Petkov I. & Schiantarelli F., 2019, "<u>Culture: Persistence and evolution</u>", Journal of Economic Growth, 24, 117–154.

Giuliano P., 2007, "Living arrangements in Western Europe: Does cultural origin matter?", Journal of the European Economic Association, 5, 927–952.

Giuliano P. & Nunn N., 2021, "<u>Understanding Cultural Persistence and Change</u>", The Review of Economic Studies, 88, 1541–1581.

Luttmer E.F.P. & Singhal M., 2011, "<u>Culture, Context, and the Taste for Redistribution</u>", American Economic Journal: Economic Policy, 3, 157–179.

- Simone Bertoli is professor CERDI, University Clermont Auvergne, and research fellow at IZA and at IC Migrations.
- + Melchior Clerc is a PhD student at CERDI, University Clermont Auvergne.
- + Jordan Loper is an associate professor at CERDI, University Clermont Auvergne.
- + Èric Roca Fernández is an assistant professor at CERDI, University Clermont Auvergne.